Theoretical Computer Science 66 (1989) 137–155
North-Holland

LOWER BOUNDS FOR SORTING OF SUMS *

Martin DIETZFELBINGER**

Lehrstuhl Informatik II, Universität Dortmund, Postfach 50 05 00, D-4600 Dortmund 50, Fed. Rep. Germany

Abstract. This paper addresses the following question: What is the complexity of sorting n numbers $x_1, \ldots, x_n$ (by comparisons), if it is known in advance that $x_1, \ldots, x_n$ are all sums of up to d out of m numbers ($n = \sum_{0 \le s \le d} \binom{m}{s}$)?

A lower bound due to Fredman concerning "Sorting $X + Y$" is extended to the following result: Let $d \ge 2$ be fixed, n, m as above. Then every comparison tree for n inputs that sorts all inputs of the form $(\sum_{r \in S} w_r \mid S \subseteq \{1, \ldots, m\}, |S| \le d)$, for $w \in \mathbb{R}^m$, has depth $\Omega(m^d) = \Omega(n)$. This lower bound is optimal. Furthermore, the case of sorting all subset sums of a vector is considered ($d = m$): Let $n = 2^m$. Then every comparison tree for n inputs that sorts all inputs of the form $(\sum_{r \in S} w_r \mid S \subseteq \{1, \ldots, m\})$, $w \in \mathbb{R}^m$, has depth $\ge 2^{\lfloor m/3 \rfloor} = \Omega(n^{1/3})$. This lower bound is exponentially larger than those previously known for this problem.

1. Introduction

One of the most thoroughly studied computational problems is the *sorting problem*: given a sequence $x = (x_1, \ldots, x_n)$ of objects from some linearly ordered domain (usually a set of numbers), determine its *ordertype*, that means, find a permutation π of $\{1, \ldots, n\}$ that satisfies $x_{\pi(1)} \le \cdots \le x_{\pi(n)}$. A natural computational model for this problem is the *comparison tree* (CT): a rooted, directed, ternary tree in which each inner node is labeled with a test "$x_i : x_j$" for some $i, j \in \{1, \ldots, n\}$, the three edges leaving such a node are labeled with "$<$", "$=$", "$>$" respectively, and each leaf is labeled with a permutation π of $\{1, \ldots, n\}$. Each instance $x = (x_1, \ldots, x_n)$ of the sorting problem determines a path in the tree (start at the root, at nodes with label "$x_i : x_j$" follow the edge with label ρ if $x_i \rho x_j$ for $\rho \in \{<, =, >\}$), and thus a leaf. The tree is said to *sort* n objects if the permutation π at the leaf determined by x satisfies $x_{\pi(1)} \le \cdots \le x_{\pi(n)}$ for every input x. As is well known, CT's that sort n inputs must have depth $\ge \log n! = n \log n + O(n)$, and there are CT's of this depth that sort n inputs.

Following Fredman [3], we consider the following generalization of the sorting problem. Let some set Γ of ordertypes be given. (That means, a set of permutations of $\{1, \ldots, n\}$; actually, one considers a family of Γ's, one for each n.) Now consider only CT's that sort all inputs x whose ordertype belongs to Γ (the leaf of the CT determined by an x with ordertype not in Γ is irrelevant). What are upper and

* Written under partial support by NSF-Grant DCR-8504247. This work is based on a part of the author's Ph.D. Thesis at the University of Illinois at Chicago, Chicago, IL, U.S.A.

** Current address: FB17, Universität - GH, Postfach 1621, D-4790, Paderborn, Fed. Rep. Germany.

lower bounds on the depth required for such CT's? In Kahn and Saks [5] this question was settled for an important class of sets Γ: there it is shown that CT's of depth $O(\log|\Gamma|)$ are sufficient if Γ is the set of all linear extensions of some partial ordering. Previously, Fredman [3] had proved an upper bound of $2n+O(\log|\Gamma|)$ on the depth necessary for an arbitrary Γ. Clearly, $\log|\Gamma|$ is always a lower bound (the "information theory lower bound"), but $\Omega(n)$ need not be a lower bound, as the example of binary search shows (here Γ is the set of ordertypes of sequences $(x_1,\ldots,x_n)$ with $x_1<\cdots<x_{n-1}$). In [3], an example is given where the summand $2n$ majorizes $\log|\Gamma|$, and actually a lower bound of $\Omega(n)$ holds, namely the problem "sort $Y+Z$", where $n=m^2$ for some m, and an input consists of all sums of the form y_i+z_j, $1\leq i,j\leq m$, for some $Y=(y_1,\ldots,y_m)$ and $Z=(z_1,\ldots,z_m)$ in $\mathbb{R}^m$.

In this paper, we further investigate the situation where the CT's are required to sort certain sums of components of $w=(w_1,\ldots,w_m)$ for $w\in\mathbb{R}^m$. The simplest case is "sorting sums of pairs":

(i) sort all sums w_r+w_s, $1\leq r<s\leq m$, for $w\in\mathbb{R}^m$. (The lower bound proof for "sorting $Y+Z$" in [3] does not carry over to this case, since there it was essential that the CT could, for example, not compare two differences $y_{j_1}-y_{j_2}$ and $y_{j_3}-y_{j_4}$. But such comparisons are admitted in our case: they correspond to tests $w_{r_1}+w_{r_4}:w_{r_2}+w_{r_3}$.)

Further, we consider "sorting sums of up to d components", where $d\geq 2$ is a constant:

(ii) sort all sums $\sum_{r\in S}w_r$, $S\subseteq\{1,\ldots,m\}$, $|S|\leq d$, for $w\in\mathbb{R}^m$.

Finally, we consider "sorting all sums of components":

(iii) sort all sums $\sum_{r\in S}w_r$, $S\subseteq\{1,\ldots,m\}$, for $w\in\mathbb{R}^m$.

More formally, for each m we fix some ordering on the subsets of $\{1,\ldots,m\}$ (once and for all). Then, for any $d\leq m$ and $n=\sum_{0\leq s\leq d}\binom{m}{s}$, sequences $(\sum_{r\in S}w_r\,|\,S\subseteq\{1,\ldots,m\},\ |S|\leq d)$ with $(w_1,\ldots,w_m)\in\mathbb{R}^m$ can naturally be regarded as elements of $\mathbb{R}^n$ (they form an m-dimensional subspace). We then consider sets Γ given as the collection of ordertypes of all sequences of the form

(i) $(\sum_{r\in S}w_r\,|\,S\subseteq\{1,\ldots,m\},\ |S|\leq 2)$, $w\in\mathbb{R}^m$ (here $n=m(m+1)/2+1$);

(ii) $(\sum_{r\in S}w_r\,|\,S\subseteq\{1,\ldots,m\},\ |S|\leq d)$, $w\in\mathbb{R}^m$, d fixed (here $n=\sum_{0\leq i\leq d}\binom{m}{i}$);

(iii) $(\sum_{r\in S}w_r\,|\,S\subseteq\{1,\ldots,m\})$, $w\in\mathbb{R}^m$ (here $n=2^m$).

In all three cases, it is easy to give upper and lower bounds on $|\Gamma|$ whose logarithms differ only by a constant factor. To be specific, we note that by a method given in [4] we may estimate $|\Gamma|$ as follows. Each possible ordertype of a sequence as in (ii) corresponds to one and only one facet of the arrangement of the hyperplanes $\{w\in\mathbb{R}^m\,|\,\sum_{r\in S}w_r=\sum_{r\in T}w_r\}$, where $S,T\subseteq\{1,\ldots,m\}$, $|S|,|T|\leq d$, in $\mathbb{R}^m$. Obviously, the number of these hyperplanes is bounded by m^{2d}. In [4] it is shown that the number of facets induced by l hyperplanes in $\mathbb{R}^m$ is $O(l^m)$. Hence $|\Gamma|=O(m^{2dm})=2^{O(m\log m)}$ in case (ii). Analogously we see that $|\Gamma|=O((3^m)^m)=2^{O(m^2)}$ in case (iii). On the other hand, we clearly have $|\Gamma|\geq m!$ in case (ii), and, as can be seen by a simple induction, $|\Gamma|\geq 2^{\Omega(m^2)}$ in case (iii). Thus, $\log|\Gamma|=\Theta(m\log m)$ in case (ii), and $\log|\Gamma|=\Theta(m^2)$ in case (iii). We see that always $\log|\Gamma|=o(n)$, so that Fredman's

result gives an upper bound of $2n+o(n)$ in all three cases. The only lower bounds that were previously known are the information theory lower bounds, which equal $\Omega(\log|\Gamma|)$ and thus are much smaller than the upper bounds.

In the following, we will show (optimal) lower bounds of $\Omega(n)$ on the depth of CT's that sove problems (i) and (ii), and a lower bound of $\Omega(n^{1/3})$ for case (iii).

The principle underlying our proofs has been widely used before (e.g., in [3, 2, 9], and before these for example in a standard proof for a lower bound for merging by comparisons): Suppose $\boldsymbol{a}=(a_1,\ldots,a_n)\in\mathbb{R}^n$ has ordertype in Γ. Suppose further that there are k distinct pairs (i_l, j_l), $l=1,\ldots,k$ (the "fooling pairs") such that $a_{i_l}<a_{j_l}$ are subsequent elements in the linear order of $\boldsymbol{a}$, and suppose that there are k other vectors $\boldsymbol{a}^l=(a_1^l,\ldots,a_n^l)\in\mathbb{R}^n$ with ordertype in Γ such that the ordertype of $\boldsymbol{a}^l$ differs from that of $\boldsymbol{a}$ only in one respect, namely that $a_{i_l}^l>a_{j_l}^l$. Then every CT that sorts inputs with ordertypes from Γ has depth $\geqslant k$. (On the path in the CT taken by $\boldsymbol{a}$ the test "$x_{i_l}:x_{j_l}$" must occur for $l=1,\ldots,k$; otherwise the tree gives an erroneous answer for at least one of $\boldsymbol{a}$ and $\boldsymbol{a}^l$.) We can think of the "difficult input" $\boldsymbol{a}$ as being the basic object to construct, in such a way that by slightly changing $\boldsymbol{a}$ we can make two of its components switch their position, but leave the order fixed otherwise— and do this in many different ways. It turns out that this general principle has to be supplemented by new ideas for constructing $\boldsymbol{a}$ to make it work in our specific context. As a basic tool in the construction we use the method of building up inputs from parts of "different orders of magnitude", or "inaccessibles", as they are called in [1].

The present results should be compared with the upper bounds from [6] for the depth of linear decision trees that recognize sets defined as unions of certain hyperplanes. (Linear decision trees (LDT's) generalize CT's in the following way: the tests at nodes in the tree are $\sum_{i=1}^n \alpha_i x_i:\alpha_0$, with $\alpha_0,\alpha_1,\ldots,\alpha_n\in\mathbb{Z}$.) It is easy to see that the contruction in [6] can be adapted in such a way that for each of the problems discussed in the present paper LDT's of depth $O(m^4\log m)$ are obtained, as long as inputs are restricted to the m-dimensional subspace

$$\left\{\left(\sum_{r\in S} w_r \;\middle|\; S\subseteq\{1,\ldots,m\},\ |S|\leqslant d\right)\;\middle|\; \boldsymbol{w}\in\mathbb{R}^m\right\}$$

respectively

$$\left\{\left(\sum_{r\in S} w_r \;\middle|\; S\subseteq\{1,\ldots,m\}\right)\;\middle|\; \boldsymbol{w}\in\mathbb{R}^m\right\}$$

of $\mathbb{R}^n$.

As we will show, CT's for these problems must have depth $\Omega(n)$. This large (in the case of "sorting all sums", exponential) difference in computing power enhances an (incomparable) result by Snir [8], which showed that some problems in $\mathbb{R}^n$ (unrestricted domain!) defined in terms of inequalities $x_i<x_j$ can be solved faster (by a constant factor) by LDT's than by CT's.

The paper is organized as follows: In Section 2 we prove the lower bound for "sorting sums of pairs", in Section 3 for "sums of up to d components", and in Section 4 for the case of "sorting arbitrary sums". (It is clear that Section 2 treats a special case of the result of Section 3, but the proof for two summands is much more transparent, so for normal expository reasons it is given separately.)

2. An optimal lower bound for sorting sums of pairs

In this section, we consider case (i) from our list.

Theorem 2.1. *Let $n = \frac{1}{2}m(m+1)+1$. Then every comparison tree for n inputs that sorts all sequences of the form*

$$\left(\sum_{r\in S} w_r \,\middle|\, S\subseteq\{1,\ldots,m\},\ |S|\leqslant 2\right),\quad \boldsymbol{w}\in\mathbb{R}^m,$$

has depth $\geqslant \lfloor m/7\rfloor^2 = \Omega(n)$.

Remark 2.2. This also holds if the CT only works for inputs with $\boldsymbol{w}\in\mathbb{N}^m$ and for $\boldsymbol{w}$'s for which all the sums $\sum_{r\in S} w_r$ are different. (The "hard" inputs we will construct will use $\boldsymbol{w}$'s consisting of positive rational numbers, and all the sums $\sum_{r\in S} w_r$ will be different. Since the problem is homogeneous, one can multiply all components by their common denominator to obtain integers.) Using a lemma from the theory of linear inequalities (see, e.g., [7, p. 319]) we see that even inputs with $\boldsymbol{w}\in\{1,2,\ldots,m^m\}^m$ are sufficient to force the lower bound. (This lemma says that if a system of inequalities $A\boldsymbol{w}\geqslant\boldsymbol{b}$ with A a $\{-1,0,1\}$-matrix and $\boldsymbol{b}$ a $\{0,1\}$-vector has a solution $\boldsymbol{w}\in\mathbb{R}^m$, then it has a solution in $\{1,2,\ldots,m^m\}^m$.)

The proof of Theorem 2.1 yields the same lower bound for the restriction of the ELEMENT DISTINCTNESS problem to inputs with the ordertypes we consider.

Corollary 2.3. *Let m, n be as in Theorem 2.1. Then every comparison tree for n inputs that, for all inputs of the form*

$$\left(\sum_{r\in S} w_r \,\middle|\, S\subseteq\{1,\ldots,m\},\ |S|\leqslant 2\right),\quad \boldsymbol{w}\in\mathbb{R}^m,$$

decides whether all components are distinct has depth $\Omega(n)$.

The proof of Theorem 2.1 also supports the following stronger version, expressed in geometrical terms.

Corollary 2.4. *There is a convex cone (i.e., a convex polytope bounded by hyperplanes through the origin) in $\mathbb{R}^m$ with $\geqslant\lfloor m/7\rfloor^2$ many facets, so that all bounding hyperplanes of the cone have the form*

$$\left\{\boldsymbol{w}\in\mathbb{R}^m \,\middle|\, \sum_{r\in S} w_r = \sum_{r\in T} w_r\right\}\quad \text{where } S,\ T\subseteq\{1,\ldots,m\},\ |S|,\ |T|\leqslant 2.$$

Proof of Theorem 2.1

The proof follows the principle described in Section 1. Let m be fixed. We construct $\boldsymbol{a} = (\sum_{r\in S} w_r \mid S \subseteq \{1, \ldots, m\}, |S| \leqslant 2)$ by defining $\boldsymbol{w} = (w_1, \ldots, w_m)$. For $p := \lfloor m/7 \rfloor$ (w.l.o.g. 7 divides m), group the m components $w_1, \ldots, w_m$ into four blocks as follows:

$$u_1, \ldots, u_{2p}, \qquad y_1, \ldots, y_p, \qquad z_1, \ldots, z_p, \qquad v_1, \ldots, v_{3p}.$$

The "fooling pairs" have the form

$$u_k + v_l : y_i + z_j, \quad 1 \leqslant i, j \leqslant p,\ k = i + j,\ l = i + 2j;$$

that is, we propose to show that for all $i, j \in \{1, \ldots, p\}$ we can produce an input $\boldsymbol{a}^{ij}$ such that its ordertype is obtained from that of $\boldsymbol{a}$ by switching the positions of these two components. $\boldsymbol{a}^{ij}$ will be obtained from $\boldsymbol{a}$ by (slightly) changing the u- and v-components in $\boldsymbol{w}$, leaving the y- and z-components fixed.

Before formally defining $\boldsymbol{w}$, $\boldsymbol{a}$, and $\boldsymbol{a}^{ij}$, we give some motivation for the way this definition is chosen.

Remark 2.5 (*Different "fooling pairs" must not be too similar*). If

$$w_{r_1} + w_{r_2} < w_{r_3} + w_{r_4} \quad \text{and} \quad w_{r_5} + w_{r_6} < w_{r_7} + w_{r_8}$$

are distinct "fooling pairs" as described in the introduction, then there can be at most one element in

$$(\{r_1, r_2\} \cap \{r_5, r_6\}) \cup (\{r_3, r_4\} \cap \{r_7, r_8\}).$$

Namely, suppose for example that $r_1 = r_5$, $r_3 = r_7$. Then the difference $(w_{r_2} + w_{r_8}) - (w_{r_4} + w_{r_6})$ changes its sign (but it should not) when $\boldsymbol{w}$ is changed to the variation $\boldsymbol{w}'$ with $w'_{r_1} + w'_{r_2} > w'_{r_3} + w'_{r_4}$ and then to the variation $\boldsymbol{w}''$ with $w''_{r_5} + w''_{r_6} > w''_{r_7} + w''_{r_8}$. Observe that our choice of the fooling pairs is such that this problem is avoided: if $u_k + v_l : y_i + z_j$ is a fooling pair, then any two of the indices i, j, k, l determine the other two.

Remark 2.6 (*A "blow-up trick"*). It is not hard to see that if $w_{r_1} + w_{r_2} < w_{r_3} + w_{r_4}$ is a "fooling pair" in $\boldsymbol{a}$, and $\boldsymbol{w}'$ is the variation of $\boldsymbol{w}$ with $w'_{r_1} + w'_{r_2} > w'_{r_3} + w'_{r_4}$, as described in the introduction, then the relation between *differences* $w_{r_5} - w_{r_6}$, $w_{r_7} - w_{r_8}$ must not change when moving from $\boldsymbol{w}$ to $\boldsymbol{w}'$ (except if these differences are $\pm(w_{r_1} - w_{r_3})$, $\pm(w_{r_4} - w_{r_2})$), since

$$(w_{r_5} - w_{r_6}) - (w_{r_7} - w_{r_8}) = (w_{r_5} + w_{r_8}) - (w_{r_6} + w_{r_7})$$

is really a difference of *sums*, which must not change its sign. The observation entails that in general we cannot make the two sums of a "fooling pair" switch their positions by just changing single components of $\boldsymbol{w}$. To overcome this difficulty, we use the following effect, which might be called a "blow-up trick".

Let $0 < \eta < \varepsilon < m^{-3}$, where $\eta/\varepsilon < m^{-3}$. Consider the two "grids" defined by $e_k := k\varepsilon$ for $1 \leq k \leq p$, and $f_k := k\varepsilon(1-(k+1)\eta)$ for $0 \leq k \leq p+1$. Clearly, the two grids interleave as follows:

$$f_1 < e_1 < f_2 < e_2 < \cdots < f_p < e_p.$$

Let $i \in \{1, \ldots, p\}$ be arbitrary. Since the differences $\varepsilon_k := f_k - f_{k-1} = \varepsilon(1-2k\eta)$ are decreasing for $k = 1, 2, \ldots, p$, it is clear that there is a "blow-up factor" b_i such that $b_i\varepsilon_i > \varepsilon > b_i\varepsilon_{i+1}$. Once such a factor b_i is fixed, it is also clear that there is an "offset" c_i such that with $f'_k := b_i f_k - c_i$, for $0 \leq k \leq p+1$, the modified (blown up, shifted) f-grid interleaves with the e-grid as follows:

$$f'_1 < e_1 < \cdots < f'_{i-1} < e_{i-1} < e_i < f'_i < f'_{i+1} < e_{i+1} < \cdots < f'_p < e_p.$$

(The only change is that $f_i < e_i$ but $f'_i > e_i$.) This is achieved without destroying the property that $f'_{k+1} - f'_k$ is decreasing. Once the requirements for b_i and c_i are clear, it is easy to find that

$$b_i := 1 + (2i+1)\eta \quad \text{and} \quad c_i := \varepsilon\eta(i^2 - m^3\eta)$$

are suitable choices. Below, this trick will be applied twofold, in a nested manner. The sums of the form $y_i + z_j$ form the (rigid) grid corresponding to the e-grid, and the u_k's and v_l's form two (flexible) grids analogous to the f-grid.

Definition of $\boldsymbol{a}$

Choose natural numbers $U < Y < Z < V$ such that all sums of up to two of them are different, excepting that $Z + Y = U + V$. Further, choose positive rational numbers ε, δ, η, μ, ν, so that ε, δ/ε, η/δ, $\mu/\eta\delta$, $\nu/\mu < m^{-3}$. Let the "perturbation terms" μ_i, ν_j (whose only purpose is to make the components of $\boldsymbol{a}$ different) be

$$\mu_i := i(i+1)\mu, \qquad \nu_j := j(j+1)\nu \qquad \text{for } 1 \leq i, j \leq p.$$

Now we can define the components of $\boldsymbol{w}$:

$$\begin{aligned}
y_i &:= Y + i(\varepsilon+\delta) + \mu_i && \text{for } 1 \leq i \leq p,\\
z_j &:= Z + j(\varepsilon+2\delta) + \nu_j && \text{for } 1 \leq j \leq p,\\
u_k &:= U + k\varepsilon(1-(k+1)\eta) && \text{for } 1 \leq k \leq 2p,\\
v_l &:= V + l\delta(1-(l+1)\eta) && \text{for } 1 \leq l \leq 3p.
\end{aligned}$$

Finally, we let

$$\boldsymbol{a} := \left(\sum_{r \in S} w_r \;\middle|\; S \subseteq \{1, \ldots, m\},\ |S| \leq 2 \right) \in \mathbb{R}^n,$$

where

$$\boldsymbol{w} = (w_1, \ldots, w_m) = (u_1, \ldots, u_{2p}, y_1, \ldots, y_p, z_1, \ldots, z_p, v_1, \ldots, v_{3p}).$$

Definition of $\boldsymbol{a}^{ij}$

For each pair $i, j \in \{1, \ldots, p\}$, we define a modified version

$$\boldsymbol{w}^{ij} = (w_1^{ij}, \ldots, w_m^{ij}) = (u_1^{ij}, \ldots, u_{2p}^{ij}, y_1, \ldots, y_p, z_1, \ldots, z_p, v_1^{ij}, \ldots, v_{3p}^{ij})$$

of $\boldsymbol{w}$. Let $k := i+j$ and $l := i+2j$. First, define "blow-up factors"

$$b_{ij} := 1+\eta(2k+1) \quad \text{and} \quad d_{ij} := 1+\eta(2l+1)$$

and "offsets"

$$c_{ij} := \eta\varepsilon k^2 - \eta^2\varepsilon m^3 \quad \text{and} \quad e_{ij} := \eta\delta l^2 - \eta^2\delta m^3.$$

Now let

$$u_{k'}^{ij} := U + b_{ij}k'\varepsilon(1-(k'+1)\eta) - c_{ij} \quad \text{for } 1 \leq k' \leq 2p,$$

$$v_{l'}^{ij} := V + d_{ij}l'\delta(1-(l'+1)\eta) - e_{ij} \quad \text{for } 1 \leq l' \leq 3p,$$

and finally

$$\boldsymbol{a}^{ij} := \left(\sum_{r \in S} w_r^{ij} \,\middle|\, S \subseteq \{1, \ldots, m\}, |S| \leq 2 \right).$$

In the remainder of this section it is shown that the inputs $\boldsymbol{a}$ and $\boldsymbol{a}^{ij}$, $1 \leq i, j \leq p$, are as required. For this, fix $i, j \in \{1, \ldots, p\}$ and let $k := i+j$ and $l := i+2j$. We must show that the only differences between the ordertypes of $\boldsymbol{a}$ and $\boldsymbol{a}^{ij}$ is that

$$y_i + z_j > u_k + v_l \quad \text{but} \quad y_i + z_j < u_k^{ij} + v_l^{ij}.$$

Let $S, T \subseteq \{1, \ldots, m\}$, $|S|, |T| \leq 2$, $S \neq T$. Define

$$\Delta := \sum_{r \in S} w_r - \sum_{r \in T} w_r \quad \text{and} \quad \Delta^{ij} := \sum_{r \in S} w_r^{ij} - \sum_{r \in T} w_r^{ij}.$$

Since we have chosen ε, δ, η, μ, and ν small enough, and since the coefficients with these numbers in the definition of the components w_r are $\leq \frac{1}{4}m^2$, we can write

$$\Delta = h_1 + h_2\varepsilon + h_3\delta + h_4\eta\varepsilon + h_5\eta\delta + h_6\mu + h_7\nu,$$

for uniquely determined integers $h_1, \ldots, h_7$, with $|h_\sigma| \leq m^2$ for $2 \leq \sigma \leq 7$. Furthermore, it is clear from the definitions that

$$|u_{k'}^{ij} - u_{k'}| \leq 2\eta\varepsilon m^2 \quad \text{for } 1 \leq k' \leq 2p, \quad \text{and}$$

$$|v_{l'}^{ij} - v_{l'}| \leq 2\eta\delta m^2 \quad \text{for } 1 \leq l' \leq 3p.$$

So if one of h_1, h_2, h_3 is nonzero, clearly, $|\Delta| > \frac{1}{2}\delta$ and $|\Delta^{ij} - \Delta| < \frac{1}{2}\delta$; hence Δ and Δ^{ij} have the same sign. Thus, we can assume w.l.o.g. that $h_1 = h_2 = h_3 = 0$; that is

$$\Delta = h_4\eta\varepsilon + h_5\eta\delta + h_6\mu + h_7\nu.$$

In particular, since $h_1 = 0$, the integer parts of the summands in $\sum_{r \in S} w_r$ and $\sum_{r \in T} w_r$ must cancel each other. This, together with the way we have chosen U, V, Y, Z, implies that one of the following two cases applies.

Case 1: The number of components of type $u_{k'}$ ($v_{l'}$, $y_{i'}$, $z_{j'}$ respectively) is the same in $\sum_{r\in S} w_r$ and $\sum_{r\in T} w_r$. (For example, $\sum_{r\in S} w_r$ is $v_{l_1}+z_{j_1}$ and $\sum_{r\in T} w_r$ is $v_{l_2}+z_{j_2}$.) This means that Δ is a sum of up to two differences of the form $u_{k_1}-u_{k_2}$, $v_{l_1}-v_{l_2}$, $y_{i_1}-y_{i_2}$, $z_{j_1}-z_{j_2}$. (The same type of difference may occur twice.)

Claim 1. $\Delta \neq 0$.

Proof. Suppose for a contradiction that $\Delta = 0$. Then $h_4 = h_5 = h_6 = h_7 = 0$. Assume for example that $\Delta = (u_{k_1} - u_{k_2}) + (w_{r_1} - w_{r_2})$ for certain k_1, k_2, r_1, r_2, where $k_1 \neq k_2$. Then $w_{r_1} - w_{r_2}$ must be of the form $u_{k_3} - u_{k_4}$, since otherwise $h_4 = (k_1(k_1+1) - k_2(k_2+1)) \neq 0$. Hence $\Delta = u_{k_1} - u_{k_2} + u_{k_3} - u_{k_4}$, for certain $k_1, \ldots, k_4$. Since $h_2 = h_3 = 0$, we have

$$k_1 - k_2 + k_3 - k_4 = 0,$$

and since $h_4 = 0$, we have

$$(k_1(k_1+1) - k_2(k_2+1)) + (k_3(k_3+1) - k_4(k_4+1)) = 0.$$

The last two equations together imply $\{k_1, k_3\} = \{k_2, k_4\}$, that is, $S = T$, a contradiction. If Δ does not contain u-components, we argue similarly for v- (or y-, or z-) components. (Here we make use of the "perturbation terms" $\mu_{i'}$ and $\nu_{j'}$.) □

Since $\Delta \neq 0$, one of h_4, h_5, h_6, h_7 is nonzero. Since Δ is a sum of differences of variables of the same type, it follows from the definition of w^{ij} that

$$\Delta^{ij} = b_{ij} h_4 \eta\varepsilon + d_{ij} h_5 \eta\delta + h_6\mu + h_7\nu,$$

hence

$$\begin{aligned}\Delta^{ij} - \Delta &= (b_{ij} - 1) h_4 \eta\varepsilon + (d_{ij} - 1) h_5 \eta\delta \\ &= (2k+1) h_4 \eta^2 \varepsilon + (2l+1) h_5 \eta^2 \delta,\end{aligned}$$

by the definition of b_{ij} and d_{ij}. The last difference has absolute value $< \Delta$ both if $h_4 \neq 0$ and if $h_4 = 0$, $h_5 \neq 0$; and if $h_4 = h_5 = 0$, we even have $\Delta = \Delta^{ij}$. So in all these cases, Δ and Δ^{ij} have the same sign.

Case 2: $\sum_{r\in S} w_r = y_{i'} + z_{j'}$ and $\sum_{r\in T} w_r = u_{k'} + v_{l'}$, for certain i', j', k', l'. (If necessary, we interchange S and T.) By our assumption, $h_2 = h_3 = 0$; hence we must have $k' = i' + j'$, $l' = i' + 2j'$. Thus

$$\Delta = \mu_{i'} + \nu_{j'} + k'(k'+1)\eta\varepsilon + l'(l'+1)\eta\delta > 0.$$

Further, $i = i'$ and $j = j'$ if and only if $k = k'$ and $l = l'$. Thus, to prove the theorem, it suffices to show the following.

Claim 2: *$\Delta^{ij} < 0$ if and only if $i' = i$ and $j' = j$.*

(Once Claim 2 is proved, we are done with the proof of Theorem 2.1: It shows that $y_i + z_j < u_k^{ij} + v_l^{ij}$, and we have seen that $y_i + z_j > u_k + v_l$. On the other hand, it shows that $y_i + z_j$, $u_k^{ij} + v_l^{ij}$ are the only components of $\boldsymbol{a}$ that switch positions if we change to $\boldsymbol{a}^{ij}$: if $\Delta^{ij} < 0$, then $i = i'$ and $j = j'$, hence $k' = k$ and $l' = l$.)

Proof. Using that $k' = i' + j'$, $l' = i' + 2j'$, write

$$\begin{aligned}\Delta^{ij} = &(\varepsilon k' + \delta l' + \mu_{i'} + \nu_{j'}) \\ &-(1+(2k+1)\eta)k'\varepsilon(1-(k'+1)\eta) + c_{ij} \\ &-(1+(2l+1)\eta)l'\delta(1-(l'+1)\eta) + e_{ij},\end{aligned}$$

and assume $\Delta^{ij} \leqslant 0$. Substituting the definitions of c_{ij} and e_{ij} and collecting terms with $\eta\varepsilon$ and $\eta\delta$ in this expression yields

$$\Delta^{ij} = [-(2k+1)k' + (k'+1)k' + k^2]\eta\varepsilon + [-(2l+1)l' + (l'+1)l' + l^2]\eta\delta + \xi,$$

for some ξ with $|\xi| < \frac{1}{2}\eta\delta$. For Δ^{ij} to be nonpositive, the coefficient of $\eta\varepsilon$ must be $\leqslant 0$. For k' varying, it assumes its minimum, which is 0, for $k = k'$. Hence this coefficient is 0. Similarly we see that the coefficient with $\eta\delta$ vanishes, and that $l = l'$. Thus, $i = 2k - l = 2k' - l' = i'$, and hence $j = j'$.

To prove the converse, assume $i = i'$, $j = j'$. This implies $k = k'$, $l = l'$, and, substituting from the definitions of c_{ij} and e_{ij},

$$\Delta^{ij} = (k+1)^2 k\eta^2\varepsilon - m^3\eta^2\varepsilon + (l+1)^2 l\eta^2\delta - m^3\eta^2\delta < 0,$$

as was to be shown. This finishes the proof of Claim 2, and the proof of Theorem 2.1. □

3. Sorting sums of a bounded number of summands

We generalize Theorem 2.1 and its proof to sums of more than two summands. The proof has the same basic structure as that of Theorem 2.1, but its greater complexity seems to justify that at least the main steps are written out in detail. Also note that here it is much more conspicuous how the components of $\boldsymbol{w}$ are built up from numbers of very different order of magnitude, thus facilitating the verification that the construction works as desired. This feature of the proof is an instance of what is called "the use of 'inaccessible' numbers" in [1].

Theorem 3.1. *Let $d \geqslant 2$, $n = \sum_{s=0}^{d} \binom{m}{s}$. Then every comparison tree for n inputs that sorts all sequences of the form*

$$\left(\sum_{r \in S} w_r \;\middle|\; S \subseteq \{1, \ldots, m\},\ |S| \leqslant d \right), \quad \boldsymbol{w} \in \mathbb{R}^m,$$

has depth $\Omega(m^d) = \Omega(n)$. (This is optimal.)

Remark 3.2. For $d \geq 3$ and m large enough (if $d \geq 5$, it is sufficient if $m \geq d^d$), a lower bound for the depth is m^d/d^{d^2}. (We will see this in the proof.) Further, Remark 2.2 and the corollaries following Theorem 2.1 apply here, mutatis mutandis, too.

Remark 3.3. The proof of Theorem 3.1 can be generalized so as to show that slightly more general decision trees—linear decision trees that can perform tests "$\sum_{r\in S} \alpha_r w_r : \sum_{r\in T} \beta_r w_r$" for $\alpha_r, \beta_r \in \mathbb{N}$ and $|S|, |T| \leq d$—still satisfy the lower bound of Theorem 3.1. (Here, the "hard" inputs depend on the size of the coefficients α_r, β_r.)

Corollary 3.4. *Let $2 \leq d < \log m/\log\log m$, for some sufficiently large m ($m \geq d^d$ is enough if $d \geq 5$). Then every comparison tree that sorts all sequences*

$$\left(\sum_{r\in S} w_r \;\middle|\; S \subseteq \{1, \ldots, m\},\ |S| \leq d \right), \quad \boldsymbol{w} \in \mathbb{R}^m,$$

has depth $\geq m^{d(1-\beta)}$ for $\beta := d \log\log m/\log m$. This lower bound is $\geq n^{1-\beta}$, where n is as in Theorem 3.1.

This follows easily from Remark 3.2.

Proof of Theorem 3.1.

We follow the principle described in Section 1. Also note that the proof parallels that of Theorem 2.1. Fix m. We construct $\boldsymbol{a} = (\sum_{r\in S} w_r \mid S \subseteq \{1, \ldots, m\}, |S| \leq d)$ by defining $\boldsymbol{w} \in \mathbb{R}^m$. Group the components $w_1, \ldots, w_m$ into $2d$ blocks as follows:

$$\begin{aligned} \boldsymbol{w} &= (w_1, \ldots, w_m) \\ &= (y_{11}, \ldots, y_{1p}, \ldots, y_{d1}, \ldots, y_{dp}, u_{11}, \ldots, u_{1,q_1}, \ldots, u_{d1}, \ldots, u_{d,q_d}). \end{aligned}$$

for certain $p, q_1, \ldots, q_d \in \mathbb{N}$ (to be defined below). We want to construct $\boldsymbol{w}$ in such a way that for every choice of $i_1, \ldots, i_d \in \{1, \ldots, p\}$ there are $k_1, \ldots, k_d$ for which

$$y_{1,i_1} + \cdots + y_{d,i_d} \quad \text{and} \quad u_{1,k_1} + \cdots + u_{1,k_d}$$

are neighbors when $\boldsymbol{a}$ is linearly ordered, and can be made switch positions by slightly altering $\boldsymbol{w}$, without causing any changes elsewhere in the order. (By the principle described in Section 1, this gives a lower bound of p^d on the depth of CT's that sort such inputs.) Again, we easily see that different "fooling pairs" must not have $\geq d$ components of $\boldsymbol{w}$ in common (cf. Remark 2.5), that is, the mapping $(i_1, \ldots, i_d) \mapsto (k_1, \ldots, k_d)$ must be such that if $(i_1, \ldots, i_d)$, $(k_1, \ldots, k_d)$ and $(i'_1, \ldots, i'_d)$, $(k'_1, \ldots, k'_d)$ are two different pairs associated with each other, then the two sequences $(i_1, \ldots, i_d, k_1, \ldots, k_d)$ and $(i'_1, \ldots, i'_d, k'_1, \ldots, k'_d)$ have $< d$ components in common. To achieve this, we let this mapping be determined by

$$(k_1, \ldots, k_d) = (i_1, \ldots, i_d)A \quad \text{for } 1 \leq i_1, \ldots, i_d \leq p,$$

where $A=(\alpha_{\sigma\tau})_{1\leqslant\sigma,\tau\leqslant d}\in\mathbb{N}^{d\times d}$ is defined by $\alpha_{\sigma\tau}:=\sigma^{\tau-1}$. (More generally, A could be any matrix with components in $\mathbb{N}$ all of whose square submatrices are regular.) We determine the sizes of the $2d$ blocks:

$$p:=\left\lfloor m\Big/\left(d+\sum_{1\leqslant\sigma,\tau\leqslant d}\alpha_{\sigma\tau}\right)\right\rfloor \quad\text{and}\quad q_\tau:=p\left(\sum_{\sigma=1}^{d}\alpha_{\sigma\tau}\right)\quad\text{for } 1\leqslant\tau\leqslant d.$$

Note. A simple calculation shows that $p\geqslant m/d^d$ for $d\geqslant 3$ and m large enough ($m\geqslant 800$ is sufficient for $d=3$, 4, and $m\geqslant d^d$ is sufficient for $d\geqslant 5$). Hence CT's that sort inputs of the kind described in the theorem have to have depth $\geqslant m^d/d^{d^2}$, if we can construct the inputs described above. For simplicity, we will assume that $d+\sum_{1\leqslant\sigma,\tau\leqslant d}\alpha_{\sigma\tau}$ divides m.

Definition of $\boldsymbol{a}$

Choose natural numbers $U_1,\ldots,U_d$, $Y_1,\ldots,Y_d$ such that all sums of up to d of these numbers (with repetitions allowed) are different, excepting that $U_1+\cdots+U_d=Y_1+\cdots Y_d$. Further, let $0<\varepsilon<m^{-2d-2}$ and define $\mu:=\varepsilon^{3d^2}$. Now let

$$y_{\sigma,i}:=Y_\sigma+i\left(\sum_{\tau=1}^{d}\alpha_{\sigma\tau}\varepsilon^\tau\right)+\mu\sum_{\rho=2}^{2d-1}i^\rho\varepsilon^{2d\sigma+\rho}\quad\text{for } 1\leqslant\sigma\leqslant d,\ 1\leqslant i\leqslant p,$$

$$u_{\tau,k}:=U_\tau+k\varepsilon^\tau-k(k+1)\varepsilon^{d+\tau}-\sum_{\rho=3}^{2d-1}k^\rho\varepsilon^{2d\tau+\rho}\quad\text{for } 1\leqslant\tau\leqslant d,\ 1\leqslant k\leqslant q_\tau.$$

Note. The "ρ-sums" are perturbation terms that are needed to make all components of $\boldsymbol{a}$ different.

Finally, we let

$$\boldsymbol{a}:=\left(\sum_{r\in S}w_r\;\middle|\;S\subseteq\{1,\ldots,m\},|S|\leqslant d\right)\in\mathbb{R}^n,$$

where $\boldsymbol{w}=(w_1,\ldots,w_m)$

$$=(y_{11},\ldots,y_{1p},\ldots,y_{d1},\ldots,y_{dp},\ u_{11},\ldots,u_{1,q_1},\ldots,u_{d1},\ldots,u_{d,q_d}).$$

Definition of $\boldsymbol{a}^{\boldsymbol{i}}$

For each $\boldsymbol{i}=(i_1,\ldots,i_d)\in\{1,\ldots,p\}^d$ we define $\boldsymbol{a}^{\boldsymbol{i}}$ as follows. First, let

$$\boldsymbol{k}:=(k_1,\ldots,k_d):=(i_1,\ldots,i_d)A.$$

Define "blow-up factors"

$$b_\tau^{\boldsymbol{i}}:=1+\varepsilon^d(2k_\tau+1),\quad 1\leqslant\tau\leqslant d,$$

and "offsets"

$$c_\tau^{\boldsymbol{i}}:=\varepsilon^{d+\tau}k_\tau^2-\varepsilon^{2d+\tau}m^{2d},\quad 1\leqslant\tau\leqslant d.$$

Then let

$$u^i_{\tau,k} := U_\tau + b^i_\tau k\varepsilon^\tau(1-(k+1)\varepsilon^d) - c^i_\tau - \sum_{\rho=3}^{2d-1} k^\rho \varepsilon^{2d\tau+\rho},$$

for $1 \leqslant \tau \leqslant d$, $1 \leqslant k \leqslant q_\tau$. Finally, let

$$\boldsymbol{w}^i := (w^i_1, \ldots, w^i_m) := (y_{11}, \ldots, y_{1p}, \ldots, y_{d1}, \ldots, y_{dp}, u^i_{11}, \ldots, u^i_{1,q_1}, \ldots, u^i_{d1}, \ldots, u^i_{d,q_d})$$

and

$$\boldsymbol{a}^i := \left(\sum_{r\in S} w^i_r \;\middle|\; S \subseteq \{1, \ldots, m\},\ |S| \leqslant d \right).$$

We must show that for each $\boldsymbol{i} = (i_1, \ldots, i_d)$ the only difference between the ordertypes of $\boldsymbol{a}$ and $\boldsymbol{a}^i$ is that

$$\sum_{\sigma=1}^{d} y_{\sigma,i_\sigma} > \sum_{\sigma=1}^{d} u_{\tau,k_\tau} \quad \text{but} \quad \sum_{\sigma=1}^{d} y_{\sigma,i_\sigma} < \sum_{\sigma=1}^{d} u^i_{\tau,k_\tau}. \tag{$*$}$$

Unfortunately, there seems to be no other way to do this than by cumbersome calculations. The principle that we will tacitly use throughout the following is that ε is so small that in the sums $\sum_{r\in S} w_r$, $\sum_{r\in T} w_r$, etc. the coefficients with different powers of ε do not interfere.

Let $S, T \subseteq \{1, \ldots, m\}$, $|S|, |T| \leqslant d$, $S \neq T$. We can assume that $S \cap T \neq \emptyset$.

Define

$$\Delta := \sum_{r\in S} w_r - \sum_{r\in T} w_r \quad \text{and} \quad \Delta^i := \sum_{r\in S} w^i_r - \sum_{r\in T} w^i_r.$$

We must show that Δ and Δ^i have different signs if and only if they are the differences of the sums in ($*$). By our definitions of the components of $\boldsymbol{w}$ and $\boldsymbol{w}^i$, we can write

$$\Delta = \sum_{\zeta=0}^{6d^2} h_\zeta \varepsilon^\zeta \quad \text{and} \quad \Delta^i = \sum_{\zeta=0}^{6d^2} h^i_\zeta \varepsilon^\zeta,$$

for uniquely determined integers h_ζ, h^i_ζ, $0 \leqslant \zeta \leqslant 6d^2$, where $|h_\zeta|, |h^i_\zeta| \leqslant m^{2d+1}$ for $1 \leqslant \zeta \leqslant 6d^2$. Since

$$|u^i_{\tau,k} - u_{\tau,k}| \leqslant m^{2d}\varepsilon^{d+\tau} \quad \text{for } 1 \leqslant \tau \leqslant d,\ 1 \leqslant k \leqslant q_\tau,$$

as immediately follows from the definitions, we have $h_\zeta = h^i_\zeta$ for $0 \leqslant \zeta \leqslant d$. So if one of these coefficients is nonzero, clearly, Δ and Δ^i have the same sign. Thus, we can assume w.l.o.g. that

$$h_\zeta = h^i_\zeta = 0 \quad \text{for } 0 \leqslant \zeta \leqslant d. \tag{1}$$

We define

$$I^S_\sigma := \{i \mid 1 \leqslant i \leqslant p,\ y_{\sigma,i} \text{ is } w_r \text{ for some } r \in S\}, \qquad 1 \leqslant \sigma \leqslant d,$$

$$J^S_\tau := \{k \mid 1 \leqslant k \leqslant q_\tau,\ u_{\tau,k} \text{ is } w_r \text{ for some } r \in S\}, \qquad 1 \leqslant \tau \leqslant d,$$

$$I^T_\sigma := \{i \mid 1 \leqslant i \leqslant p,\ y_{\sigma,i} \text{ is } w_r \text{ for some } r \in T\}, \qquad 1 \leqslant \sigma \leqslant d,$$

$$J^T_\tau := \{k \mid 1 \leqslant k \leqslant q_\tau,\ u_{\tau,k} \text{ is } w_r \text{ for some } r \in T\}, \qquad 1 \leqslant \tau \leqslant d.$$

Since $S \cap T = \emptyset$, we have $I_\sigma^S \cap I_\sigma^T = \emptyset$ and $J_\tau^S \cap J_\tau^T = \emptyset$. With these definitions, Δ and Δ^i can be written as follows:

$$\Delta = \sum_{\sigma=1}^{d} \left(\sum_{i \in I_\sigma^S} y_{\sigma,i} - \sum_{i \in I_\sigma^T} y_{\sigma,i} \right) + \sum_{\tau=1}^{d} \left(\sum_{k \in J_\tau^S} u_{\tau,k} - \sum_{k \in J_\tau^T} u_{\tau,k} \right), \tag{2}$$

$$\Delta^i = \sum_{\sigma=1}^{d} \left(\sum_{i \in I_\sigma^S} y_{\sigma,i} - \sum_{i \in I_\sigma^T} y_{\sigma,i} \right) + \sum_{\tau=1}^{d} \left(\sum_{k \in J_\tau^S} u^i_{\tau,k} - \sum_{k \in J_\tau^T} u^i_{\tau,k} \right). \tag{2'}$$

Equation (1) means that the coefficients with ε^ζ in the summands of $\sum_{r \in S} w_r$ and $\sum_{r \in T} w_r$ cancel each other for $0 \leqslant \zeta \leqslant d$. For $\zeta = 0$ this means

$$\sum_{\sigma=1}^{d} (|I_\sigma^S| - |I_\sigma^T|) Y_\sigma + \sum_{\tau=1}^{d} (|J_\tau^S| - |J_\tau^T|) U_\tau = 0, \tag{3}$$

and for $1 \leqslant \zeta \leqslant d$ we get, using the definitions of $y_{\sigma,i}$ and $u_{\tau,k}$:

$$\sum_{\sigma=1}^{d} \left(\sum_{i \in I_\sigma^S} i - \sum_{i \in I_\sigma^T} i \right) \alpha_{\sigma\tau} + \left(\sum_{k \in J_\tau^S} k - \sum_{k \in J_\tau^T} k \right) = 0 \quad \text{for } 1 \leqslant \tau \leqslant d. \tag{4}$$

Equation (3) asserts that two sums of $\leqslant d$ summands out of $U_1, \ldots, U_d, Y_1, \ldots, Y_d$ are equal. Hence, by the way these integers were chosen, one of the following two cases applies.

Case 1: The equality is trivial, that is,

$$|I_\sigma^S| = |I_\sigma^T| \quad \text{for } 1 \leqslant \sigma \leqslant d, \quad \text{and} \quad |J_\tau^S| = |J_\tau^T| \quad \text{for } 1 \leqslant \tau \leqslant d.$$

Thus, in (4) the differences $\sum_{i \in I_\sigma^S} i - \sum_{i \in I_\sigma^T} i$ for $I_\sigma^S = \emptyset$ and $\sum_{k \in J_\tau^S} k - \sum_{k \in J_\tau^T} k$ for $J_\tau^S = \emptyset$ all vanish. That is, (4) asserts that the vector consisting of the d numbers

$$\sum_{i \in I_\sigma^S} i - \sum_{i \in I_\sigma^T} i, \quad I_\sigma^S \neq \emptyset, \qquad \sum_{k \in J_\tau^S} k - \sum_{k \in J_\tau^T} k, \quad J_\tau^S \neq \emptyset,$$

and $\quad d - |\{\sigma | I_\sigma^S \neq \emptyset\}| - |\{\tau | J_\tau^S \neq \emptyset\}|$ many 0's

solves a homogeneous system of equations whose matrix is obtained from A by replacing the σ-th row by e_σ, the σ-th unit vector, for all σ with $I_\sigma^S = \emptyset$. But this matrix is regular, since all square submatrices of A are, by choice of A. Hence the solution to this sytem is trivial, and we get

$$\sum_{i \in I_\sigma^S} i = \sum_{i \in I_\sigma^T} i \quad \text{for } 1 \leqslant \sigma \leqslant d, \quad \text{and} \quad \sum_{k \in J_\tau^S} k = \sum_{k \in J_\tau^T} k \quad \text{for } 1 \leqslant \tau \leqslant d. \tag{5}$$

Since $|J_\tau^S| = |J_\tau^T|$ for $1 \leqslant \tau \leqslant d$, all the "offsets" occurring in Δ^i cancel each other. More precisely, we get from (2), (2'), (5), and the definitions of $y_{\sigma,i}$ and $u_{\tau,k}$ that

$$\begin{aligned} \Delta - \Delta^i &= \sum_{\tau=1}^{d} (b_\tau^i - 1) \left(\sum_{k \in J_\tau^S} k(k+1) - \sum_{k \in J_\tau^T} k(k+1) \right) \varepsilon^{\tau+d} \\ &= \sum_{\tau=1}^{d} (2k_\tau + 1) \left(\sum_{k \in J_\tau^S} k(k+1) - \sum_{k \in J_\tau^T} k(k+1) \right) \varepsilon^{\tau+2d} \\ &= \sum_{\tau=1}^{d} (2k_\tau + 1) h_{d+\tau} \varepsilon^{\tau+2d}. \end{aligned}$$

If $h_\zeta \neq 0$ for some $\zeta \in \{d+1, \ldots, 2d\}$, let $d+\tau_0$ be the smallest such ζ. Then from (1) it follows that Δ and $h_{d+\tau_0}$ have the same sign (in particular, $\Delta \neq 0$), and from the last equation it follows that

$$|\Delta^i - \Delta| < |(2k_\tau + 2)h_{d+\tau_0}\varepsilon^{\tau+2d}| < |\Delta|,$$

so Δ^i and Δ have the same sign. Thus, we can assume from here on that

$$h_\zeta = 0 \quad \text{for } 0 \leq \zeta \leq 2d. \tag{6}$$

This immediately implies that $\Delta = \Delta^i$. All that remains to be shown to finish Case 1 is the following.

Claim 1. $\Delta \neq 0$.

Proof. Suppose for a contradiction that $\Delta = 0$, that is, that $h_\zeta = 0$ for $0 \leq \zeta \leq 6d^2$. Substituting (5) and (6) into (2), together with the definition of $y_{\sigma,i}$ and $u_{\tau,k}$, yields

$$\begin{aligned} 0 = \Delta \\ = \mu \sum_{\sigma=1}^{d} \sum_{\rho=2}^{2d-1} \left(\sum_{i \in I_\sigma^S} i^\rho - \sum_{i \in I_\sigma^T} i^\rho \right) \varepsilon^{2d\sigma+\rho} \\ + \sum_{\tau=1}^{d} \sum_{\rho=3}^{2d-1} \left(\sum_{k \in J_\tau^S} k^\rho - \sum_{k \in J_\tau^T} k^\rho \right) \varepsilon^{2d\tau+\rho}. \end{aligned}$$

All coefficients with different powers of ε vanish. That is,

$$\sum_{i \in I_\sigma^S} i^\rho - \sum_{i \in I_\sigma^T} i^\rho = 0 \quad \text{for } 2 \leq \rho < 2d,\ 1 \leq \sigma \leq d, \tag{7}$$

$$\sum_{k \in J_\tau^S} k^\rho - \sum_{k \in J_\tau^T} k^\rho = 0 \quad \text{for } 3 \leq \rho < 2d,\ 1 \leq \tau \leq d. \tag{8}$$

Suppose (for a contradiction) that $J_\tau^S \neq \emptyset$ for some τ. Equations (5), (6), and the assumption for Case 1 together imply that the equality in (8) actually holds for *all* ρ, $0 \leq \rho < 2d$. This means that the matrix whose $2|J_\tau^S|$ columns consist of the powers k^ρ $(0 \leq \rho \leq 2d)$ for $k \in J_\tau^S \cup J_\tau^T$ (note that these are all different integers) has rank $< 2|J_\tau^S|$. But this cannot be the case, since for example the first $2|J_\tau^S|$ rows of this matrix are linearly independent. We conclude that $J_\tau^S = \emptyset$ for $1 \leq \tau \leq d$.

Similarly, we get from (7), (5), and the assumption for Case 1 that $I_\sigma^S = \emptyset$ for $1 \leq \sigma \leq d$. Thus, we have shown that $S = \emptyset$, hence (by the assumption for Case 1) $T = \emptyset$, which contradicts our initial assumption $S \neq T$. This finishes the proof of Claim 1, and thus Case 1. □

Case 2. Equation (3) is really $U_1 + \cdots + U_d = Y_1 + \cdots + Y_d$; that is,

$$|I_\sigma^S| = |J_\tau^T| = 1 \quad \text{and} \quad I_\sigma^T = J_\tau^S = \emptyset \quad \text{for } 1 \leq \sigma, \tau \leq d \text{ (or vice versa).}$$

Let i'_σ be the unique element of I^S_σ for $1 \leq \sigma \leq d$, and let k'_τ be the unique element of J^T_τ for $1 \leq \tau \leq d$. Then (2) turns into

$$\Delta = \sum_{\sigma=1}^{d} y_{\sigma,i'_\sigma} - \sum_{\tau=1}^{d} u_{\tau,k'_\tau}, \tag{9}$$

and (2′) becomes

$$\Delta^i = \sum_{\sigma=1}^{d} y_{\sigma,i'_\sigma} - \sum_{\tau=1}^{d} u^i_{\tau,k'_\tau}. \tag{9′}$$

Equation (4) turns into

$$\sum_{\sigma=1}^{d} i'_\sigma \alpha_{\sigma\tau} = k'_\tau \quad \text{for } 1 \leq \tau \leq d, \tag{10}$$

which is just

$$(i'_1, \ldots, i'_d)A = (k'_1, \ldots, k'_d). \tag{10′}$$

This implies, by the regularity of A and the definition of $\boldsymbol{k}$:

$$i_\sigma = i'_\sigma \text{ for } 1 \leq \sigma \leq d \quad \text{if and only if} \quad k_\tau = k'_\tau \text{ for } 1 \leq \tau \leq d. \tag{11}$$

Substituting from the definitions into (9) and canceling according to (10) yields

$$\Delta = \sum_{\sigma=1}^{d} \mu \sum_{\rho=2}^{2d-1} i'^{\rho}_\sigma \varepsilon^{2d\sigma+\rho} + \sum_{\tau=1}^{d} k'_\tau(k'_\tau+1)\varepsilon^{d+\tau} + \sum_{\rho=3}^{2d-1} k'^{\rho}_\tau \varepsilon^{2d\tau+\rho} > 0.$$

Thus, taking (11) into account, it is enough to prove the following.

Claim 2. *$\Delta^i < 0$ if and only if $(i_1, \ldots, i_d) = (i'_1, \ldots, i'_d)$.*

Proof. Assume first that $\Delta^i \leq 0$. Then it is clear (from (1)) that $h^i_{d+1} \leq 0$. But

$$h^i_{d+1} = -(2k_1+1)k'_1 + (k'_1+1)k'_1 + k_1^2,$$

regarded as a function of k'_1, attains its minimum, which is 0, for $k'_1 = k_1$. Hence $h^i_{d+1} = 0$, and h^i_{d+2} must be nonpositive. Repeating this argument, we find that $h^i_\zeta = 0$ for $d+1 \leq \zeta \leq 2d$, and that $k_\tau = k'_\tau$ for $1 \leq \tau \leq d$. Applying (11), we get $i_\sigma = i'_\sigma$ for $1 \leq \sigma \leq d$.

For the converse, assume that $(i_1, \ldots, i_d) = (i'_1, \ldots, i'_d)$. This implies $(k_1, \ldots, k_d) = (k'_1, \ldots, k'_d)$, and

$$h^i_{\tau+d} = -(2k_\tau+1)k'_\tau + (k'_\tau+1) + k_\tau^2 = 0 \quad \text{for } 1 \leq \tau \leq d.$$

Finally, substituting the definitions of $u^i_{\tau,k}$ and $y_{\sigma,i}$ into (9′), and simplifying, we obtain

$$\begin{aligned}\Delta^i = &\sum_{\sigma=1}^{d} \mu \sum_{\rho=2}^{2d-1} i^\rho \varepsilon^{2d\sigma+\rho} + \sum_{\tau=1}^{d} \sum_{\rho=3}^{2d-1} k_\tau^\rho \varepsilon^{2d\tau+\rho} \\ &+ \sum_{\tau=1}^{d} (\varepsilon^{d+\tau} k_\tau^2 - \varepsilon^{2d+\tau} m^{2d^2}) - \sum_{\tau=1}^{d} k_\tau^2 \varepsilon^{d+\tau} \\ &+ \sum_{\tau=1}^{d} k_\tau \varepsilon^{d+\tau} \varepsilon^d (2k_\tau+1) < 0,\end{aligned}$$

as was to be shown. This finishes Case 2 and the proof of Theorem 3.1. □

4. A lower bound for sorting arbitrary sums

Theorem 4.1. *Let $n = 2^m$. Then every comparison tree for inputs from $\mathbb{R}^n$ that sorts all sequences of the form*

$$\left(\sum_{r \in S} w_r \,\middle|\, S \subseteq \{1, \ldots, m\} \right), \quad w \in \mathbb{R}^m,$$

has depth $\geq 2^{\lfloor m/3 \rfloor} = \Omega(n^{1/3})$.

Note. The two corollaries following Theorem 2.1 apply here, mutatis mutandis, too.

Proof. When phrased in geometrical terms, the principle for our lower bound proofs formulated in Section 1 says that it suffices to show that for every m the following is true:

(A_m) There are points $w = (w_1, \ldots, w_m)$ and $w^l = (w_1^l, \ldots, w_1^l)$, $1 \leq l \leq 2^{m/3}$, in $\mathbb{R}^m$, none of them lying on a hyperplane of the form

$$\left\{ v \in \mathbb{R}^m \,\middle|\, \sum_{r \in S} v_r = \sum_{r \in T} v_r \right\} \quad \text{where } S, T \subseteq \{1, \ldots, m\}, \tag{*}$$

such that for every l the straight line from w to w^l intersects exactly one hyperplane (called H_l) of type (*), and $H_l \neq H_{l'}$ for $1 \leq l \neq l' \leq 2^{m/3}$.

(Note that the sets S, T in (*) are uniquely determined by the hyperplane if we demand that $S \cap T = \emptyset$ and $\sum_{r \in T} w_r < \sum_{r \in S} w_r$. In this case, for every U with $U \cap (S \cup T) = \emptyset$, the elements $\sum_{r \in S \cup U} v_r$ and $\sum_{r \in T \cup U} v_r$ will switch their positions in the sequence $(\sum_{r \in S} v_r \,|\, S \subseteq \{1, \ldots, m\})$, when v crosses the hyperplane (*).)

We prove (A_m) by induction on m, where $3 | m$. The case $m = 3$ is trivial. So assume $m \geq 3$, and fix points w and w^l and hyperplanes H_l ($1 \leq l \leq 2^{m/3}$) as described in (A_m). For each l, fix the (unique) disjoint sets S_l, $T_l \subseteq \{1, \ldots, m\}$ with

$$H_l = \left\{ v \in \mathbb{R}^m \,\middle|\, \sum_{r \in S_l} v_r = \sum_{r \in T_l} v_r \right\} \quad \text{and} \quad \sum_{r \in T_l} w_r < \sum_{r \in S_l} w_r.$$

We choose $\varepsilon > 0$ so small that for all l, for all v on the straight line from w to w^l, and for all S, $T \subseteq \{1, \ldots, m\}$ with $\{v \in \mathbb{R}^m \,|\, \sum_{r \in S} v_r = \sum_{r \in T} v_r\} \neq H_l$ we have

$$\left| \sum_{r \in S} v_r - \sum_{r \in T} v_r \right| \geq \varepsilon.$$

Choose $w_{m+2} > w_{m+1} > \sum_{1 \leq r \leq m} w_r$ with $w_{m+2} - w_{m+1} < \varepsilon$. Further, choose $w_{m+3} > w_{m+1} + w_{m+2}$ with $w_{m+3} - (w_{m+1} + w_{m+2}) < \varepsilon$. Observe that for arbitrary S, $T \subseteq \{1, \ldots, m\}$ with $\sum_{r \in T} w_r < \sum_{r \in S} w_r$ the sums formed by these together with

$w_{m+1}, w_{m+2}, w_{m+3}$ are arranged as follows:

$$\begin{aligned}\sum_{r\in T} w_r &< \sum_{r\in S} w_r < w_{m+1} + \sum_{r\in T} w_r < w_{m+2} + \sum_{r\in T} w_r \\ &< w_{m+1} + \sum_{r\in S} w_r < w_{m+2} + \sum_{r\in S} w_r < w_{m+1} + w_{m+2} + \sum_{r\in T} w_r \\ &< w_{m+3} + \sum_{r\in T} w_r < w_{m+1} + w_{m+2} + \sum_{r\in S} w_r < w_{m+3} + \sum_{r\in S} w_r \\ &< w_{m+1} + w_{m+3} + \sum_{r\in T} w_r, \end{aligned} \tag{12}$$

by the choice of ε and w_{m+1}, w_{m+2}, w_{m+3}.

In the following, we show how for each l we can change $(w_1, \ldots, w_{m+3})$ in two different ways, each time crossing exactly one hyperplane of the type

$$\left\{ \boldsymbol{u} \in \mathbb{R}^{m+3} \,\middle|\, \sum_{r\in S} u_r = \sum_{r\in T} u_r \right\}, \tag{**}$$

so that different hyperplanes are crossed for different l. Fix l.

We describe the first change in detail. Move in $\mathbb{R}^{m+3}$ along straight lines as follows: Start from $(w_1, \ldots, w_{m+3})$. Leaving w_{m+1} fixed, change w_{m+2} and w_{m+3} to new numbers w'_{m+2} and w'_{m+3} such that

$$0 < w'_{m+2} - w_{m+1} < w_{m+3} - (w_{m+1} + w_{m+2}) = w'_{m+3} - (w_{m+1} + w'_{m+2}) < \varepsilon.$$

Let $\boldsymbol{w}' := (w_1, \ldots, w_m, w_{m+1}, w'_{m+2}, w'_{m+3})$. Observe that $\boldsymbol{w}'$ also satisfies (12), hence no hyperplane of type (**) separates $\boldsymbol{w}$ and $\boldsymbol{w}'$. Then, leaving w_{m+1}, w'_{m+2}, w'_{m+3} fixed, move along the line from $\boldsymbol{w}$ to $\boldsymbol{w}^l$ in the first m components of $(w_1, \ldots, w_m, w_{m+1}, w'_{m+2}, w'_{m+3})$. If S and T are such that $H_l \neq \{\boldsymbol{v} \in \mathbb{R}^m \mid \sum_{r\in S} v_r = \sum_{r\in T} v_r\}$, then $(v_1, \ldots, v_m, w_{m+1}, w'_{m+2}, w'_{m+3})$ will satisfy (12) for all $\boldsymbol{v}$ on the line from $\boldsymbol{w}$ to $\boldsymbol{w}^l$, since, by the choice of ε,

$$\sum_{r\in S} v_r - \sum_{r\in T} v_r > \varepsilon > w'_{m+3} - (w_{m+1} + w'_{m+2}) > w'_{m+2} - w_{m+1}$$

for all these $\boldsymbol{v}$. However,

$$\begin{aligned}\sum_{r\in S_l} w_r - \sum_{r\in T_l} w_r &> w'_{m+3} - (w_{m+1} + w'_{m+2}) > w'_{m+2} - w_{m+1} \\ &> 0 > \sum_{r\in S_l} w^l_r - \sum_{r\in T_l} w^l_r,\end{aligned}$$

so for some $\boldsymbol{v}$ on the straight line from $\boldsymbol{w}$ to $\boldsymbol{w}^l$ we have

$$w'_{m+3} - (w_{m+1} + w'_{m+2}) > \sum_{r\in S_l} v_r - \sum_{r\in T_l} v_r > w'_{m+2} - w_{m+1},$$

that is

$$w_{m+1} + \sum_{r\in T_l} v_r < m'_{m+2} + \sum_{r\in T_l} v_r < w_{m+1} + \sum_{r\in S_l} v_r < w'_{m+2} + \sum_{r\in S_l} v_r,$$

$$w_{m+1} + w'_{m+2} + \sum_{r\in T_l} v_r < w_{m+1} + w'_{m+2} + \sum_{r\in S_l} v_r < w'_{m+3} + \sum_{r\in T_l} v_r < w'_{m+3} + \sum_{r\in S_l} v_r,$$

that is, w' and $(v_1, \ldots, v_m, w_{m+1}, w'_{m+2}, w'_{m+3})$ are separated by the hyperplane

$$\left\{ u \in \mathbb{R}^{m+3} \;\middle|\; \sum_{r \in S_l} u_r + u_{m+1} + u_{m+2} = \sum_{r \in T_l} u_r + u_{m+3} \right\}.$$

We let $w^{l1} := (v_1, \ldots, v_m, w_{m+1}, w'_{m+2}, w'_{m+3})$.

To find a second variation of $(w_1, \ldots, w_{m+3})$ separated from this point by only one hyperplane of type (**), we start from $(w_1, \ldots, w_{m+3})$ again, and, leaving w_{m+1} and w_{m+2} fixed, change w_{m+3} to w''_{m+3} in such a way that

$$0 < w''_{m+3} - (w_{m+1} - w_{m+2}) < w_{m+2} - w_{m+1} < \varepsilon.$$

Let $\bar{w}'' := (w_1, \ldots, w_{m+2}, w''_{m+3})$. Now, just as before, leave the last three components fixed and move in the first m components on the straight line from w to w^l. Arguing as above, we see that the first hyperplane of type (**) crossed on this path is

$$\left\{ u \in \mathbb{R}^{m+3} \;\middle|\; \sum_{r \in S_l} u_r + u_{m+1} = \sum_{r \in T_l} u_r + u_{m+2} \right\}.$$

So there is a w^{l2} that is separated from w only by this hyperplane. Since this can be done for every l, we can actually cross $2 \cdot 2^{m/3}$ different hyperplanes of type (**), by slightly altering $(w_1, \ldots, w_{m+3})$. This finishes the induction step. □

5. Concluding remarks

Although Theorem 3.1 can be generalized to slightly more powerful decision trees (see Remark 3.3), the proof method will not work for linear decision trees whose tests involve more than $2d$ variables. But it seems possible that still a lower bound larger than the information theory bound holds. Similarly, it is open if the lower bound of Theorem 4.1 (sorting arbitrary sums) holds for linear decision trees that can perform tests like "$\sum_{1 \le r \le m} \alpha_r w_r : \alpha_0$" for certain (small) $\alpha_0, \alpha_1, \ldots, \alpha_m \in \mathbb{Z}$. Also, it would be of interest to show an exponential (in m) lower bound for the Knapsack problem "is $w \in \{w \in \mathbb{R}^m \mid \sum_{r \in S} w_r = 1 \text{ for some } S \subseteq \{1, \ldots, m\}\}$?" on decision trees that can use tests "$\sum_{r \in S} w_r : \sum_{r \in T} w_r$" and "$\sum_{r \in S} w_r : 1$?". In [9] and in [1] some partial results in this direction are given, but it seems that in all cases new techniques have to be found in order to find complete answers. As mentioned before, there is a limit to such generalizations: using arbitrary linear tests with integer coefficients, linear decision trees can solve all the problems discussed in this section in depth $O(m^4 \log m)$ [6]. It might be interesting to try to find the boundary (in terms of the size of the coefficients) where our lower bounds yield to the upper bound from [6].

Acknowledgment

I would like to thank W. Maass, U. Peled, and G. Turán for helpful discussions, and F. Meyer auf der Heide for pointing out Snir's result.

References

[1] M. Dietzfelbinger and W. Maass, Lower bound arguments with "inaccessible" numbers, *J. Comput. System Sci.* **36** (1988) 313-335.

[2] D.P. Dobkin and R.J. Lipton, On the complexity of computations under varying sets of primitives, *J. Comput. System Sci.* **18** (1979) 86-91.

[3] M.L. Fredman, How good is the information theory bound in sorting?, *Theoret. Comput. Sci.* **1** (1976) 355-361.

[4] L.H. Harper, T.H. Payne, J.E. Savage and E. Strauss, Sorting $X + Y$, *Comm. Assoc. Comput. Mach.* **18** (1975) 347-349.

[5] J. Kahn and M. Saks, Balancing poset extensions, *Order* **1** (1984) 113-126.

[6] F. Meyer auf der Heide, A polynomial linear search algorithm for the n-dimensional knapsack problem, *J. Assoc. Comput. Mach.* **31** (1984) 668-676.

[7] Ch.H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity* (Prentice-Hall, Englewood Cliffs, NJ, 1982).

[8] M. Snir, Comparisons between linear functions can help, *Theoret. Comput. Sci.* **19** (1982) 321-330.

[9] E. Ukkonen, Exponential lower bounds for some NP-complete problems in a restricted linear decision tree model, *BIT* **23** (1983) 181-192.